An NMR study on the β-hairpin region of barnase

José L Neira and Alan R Fersht

Background: The β-hairpin of barnase (residues Ser92–Leu95) has been proposed in theoretical and protein engineering studies to be an initiation site for folding [1]. There is evidence for residual structure in this region from NMR studies of the denatured protein under different denaturing conditions [2,3]. A more detailed analysis is possible by NMR studies of isolated fragments.

Results: Protons of fragments B(80–110) and B(69–110) in 6 M urea have non-random chemical shifts. Non-native long-range and medium-range NOE contacts with the aromatic moiety of Trp94 indicate that it is involved in a β-turn-like or α-helix-like conformation. Also, the sidechains of Trp71, Tyr79, Phe82, Tyr90, Tyr97, His102, Tyr103 and Phe106 show non-native hydrophobic contacts. Non-random conformational shifts and sequential NN(*i*, *i*+1) NOE contacts are clustered to one of the β-strands and one of the loop regions.

Conclusions: The hairpin region of barnase adopts β-turn-like or α-helix-like conformations, which are weakly populated even in 6 M urea. The hairpin region is a potential nucleation site in folding that may consolidate on docking with the first α-helix. The other residues that have conformational preferences form a β-strand and one of the loop regions in the native intact protein, but they do not constitute a nucleation site.

Address: MRC Unit for Protein Function and Design, Cambridge Centre for Protein Engineering, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK.

Correspondence: Alan R Fersht
e-mail: arf10@cam.ac.uk

Key words: nucleation site, protein folding, tertiary interactions, unfolded states

Received: 06 Mar 1996
Revisions requested: 26 Mar 1996
Revisions received: 11 Apr 1996
Accepted: 02 May 1996

Published: 31 May 1996
Electronic identifier: 1359-0278-001-00231

Folding & Design 31 May 1996, 1:231–241

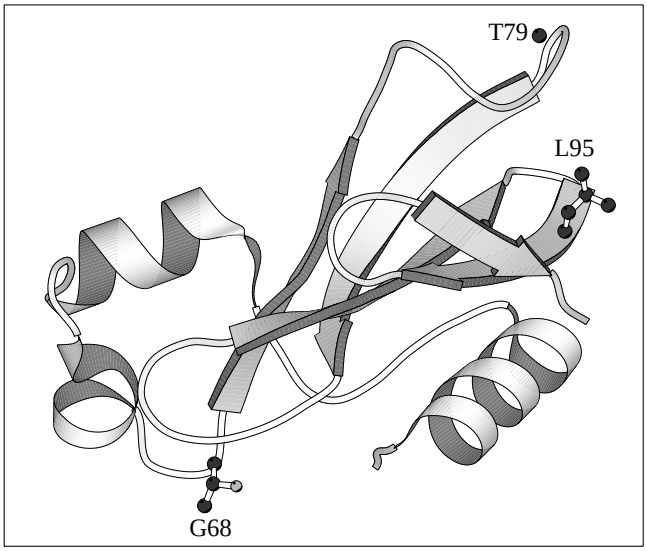
Introduction

A complete resolution of a folding pathway requires the description of all states along the reaction coordinate: the starting point (the denatured state); all the intermediate states; the high energy transition states; and the final folded state. The conformations of folded states can be characterized by X-ray and NMR techniques. Intermediates states appearing late in the folding reaction can be described by NMR [4] and the protein engineering method [1,5]. The latter technique is the only experimental method for mapping transition states at the level of individual residues and has been applied to barnase, chymotrypsin inhibitor 2 (see [6] for a review) and chemotactic protein (CheY) [7]. In contrast, little is known about the early events that precede the formation of those intermediates and the main transition state. It is important to understand the nature of the early states for a complete and detailed knowledge of the folding pathway.

It has been suggested that folding initiates by formation of nucleation sites, involving residues very close in the structure [8,9]. It has been proposed that nucleation sites are weakly populated in the denatured states, but they develop in the transition state on making long-range interactions [10,11], a result supported by lattice simulations [12]. Hydrophobic burial has been postulated to drive the formation of those sites [13], but only when there are a sufficient number of hydrophobic residues involved [14]. There are now several NMR studies of denatured proteins which all find evidence of clustering around hydrophobic residues but with few, if any, tertiary interactions [15].

Barnase is the extracellular ribonuclease produced by the prokaryote *Bacillus amyloliquefaciens*. It is a small monomeric (110 residues, 12382 Da) α + β single domain enzyme with three α-helices in the first 45 residues and five antiparallel β-strands in the last 65 residues (Fig. 1). The elements of secondary structure are the following: the N-terminal region (residues 1–5); $\alpha\text{-helix}_1$ (residues 6–18); loop_1 (residues 19–25); $\alpha\text{-helix}_2$ (residues 26–34); loop_2 (residues 35–40); $\alpha\text{-helix}_3$ (residues 41–46); β-turn (residues 47–49); $\beta\text{-strand}_1$ (residues 50–55); loop_3 (residues 56–69); $\beta\text{-strand}_2$ (residues 70–76); loop_4 (residues 77–84); $\beta\text{-strand}_3$ (residues 85–91); β-turn (residues 92–95); $\beta\text{-strand}_4$ (residues 96–99); loop_5 (residues 100–105); $\beta\text{-strand}_5$ (residues 106–108); and C-terminal residues (residues 109–110). The structures of an intermediate and the transition state of barnase folding have been characterized by a combination of the protein engineering method and NMR [16–19]. The structures of both species resemble a distorted form of the native state, with the two α-helices formed in their central and C-terminal region. The protein engineering procedure used to describe those states has been discussed extensively elsewhere [1,6,7]. Briefly, site-directed mutagenesis is used to remove a suitable sidechain interaction. The resulting mutant is characterized by equilibrium denaturation experiments (changes in stability upon mutation) and by kinetic unfolding and refolding experiments (measure-

Figure 1



General overview of barnase, produced with MOLSCRIPT [52], indicating the mutated residues.

ment of changes in activation energies and energy levels of transition state or any intermediate). We can describe the extent of the interaction of each residue at different states of the folding pathway by the ratio of the change in energy level of a transition state or a folding intermediate on mutation to the change in the overall free energy of folding. This gives a series of values of a parameter Φ. A value of $\Phi = 0$ implies that the structure at the site of mutation (in the transition or in the intermediate state) is as unfolded as it is in the unfolded state. In contrast, $\Phi = 1$ indicates that the structure at the site of the mutation is as folded as it is in the folded state. The fractional values result from incomplete formation of structure due either to partial non-covalent bond formation or parallel folding pathways.

Studies of denatured states may be complemented by NMR analysis of protein fragments. In barnase, using the protein engineering method, we have shown that fragments comprising the α-helical N-terminal sequence have very little helical structure [20,21] as does the denatured state of the intact protein [2,3]. There is evidence of structure in the denatured state in the last β-turn of the protein (residues Ser92–Leu95) [2,3]. It has been proposed that this region constitutes an initiation site for folding [1]. Here, we study several C-terminal fragments which comprise the β-turn region. In order to produce very large fragments, we have used the protein engineering procedure developed on barnase [20] and chymotrypsin inhibitor 2 [22,23] of introducing a methionine at a specific position of the polypeptide chain followed by cleavage with cyanogen bromide. Our studies show the presence of a hydrophobic clustering centred around Trp94, even in the presence of high urea concentrations. These results, with previous protein engineering studies, suggest that this region may act as an initiation folding site in the very early events of barnase folding.

Results

In order to identify the presence of residual structure in every fragment, we have used a combined strategy of analyzing the chemical shifts, the temperature coefficients and NOESY data. We have not used the coupling constant, $^3J_{NHH\alpha}$, mainly because signal overlapping precludes its measurement (only those of B(96–110) could be measured for all the residues, which were all random coil values).

B(69–110) and B(80–110) fragments

B(80–110) corresponds to half of $loop_4$, β-$strand_3$, the β-turn region, β-$strand_4$, $loop_5$, β-$strand_5$ and the last two residues of native barnase; B(69–110) comprises the same elements of secondary structure, β-$strand_2$ and the complete $loop_4$. Both fragments are highly insoluble in water, but can be studied in 6 M urea. Sample aggregation was tested by measuring the linewidths and chemical shifts in 1D-NMR experiments for every peptide at different concentrations. No changes were observed in the concentration range examined (0.25–2.5 mM). Although there are large overlaps of the signals under these conditions (especially in B(69–110)), the fragments are short enough to provide an almost complete assignment of the resonances. The assignments of both fragments are provided as Supplementary material. Selected regions of a NOESY experiment on B(80–110) are shown in Figure 2.

Chemical shift analysis

The factors affecting chemical shifts are numerous and they are not fully understood. However, semi-empirical and theoretical approaches have indicated the structural information inherent in chemical shift values [24,25]. Significant deviations in chemical shifts from values for model compounds provide preliminary evidence for 'non-random' conformations [26]. Here, we have used the accepted ranges for the random coil outside which deviations are considered to be significant: $\Delta\delta > 0.2$ ppm for labile protons and $\Delta\delta > 0.1$ ppm for non-labile protons [2,3,27].

In B(69–110) (Fig. 3a) and B(80–110) (Fig. 3b), the conformational shifts for the H_α protons are nearly the same in the common regions for both fragments, i.e. in half of $loop_4$, β-$strand_4$, the β-turn region, β-$strand_4$, $loop_5$, β-$strand_5$ and the last two residues. Small differences (on average ± 0.05 ppm) are observed for the H_α protons of residues between Thr99, Asp101 and Thr105. There are several possible explanations for these H_α chemical shift differences. Subtle changes in the pH and urea concentra-

Figure 2

NOESY spectra: selected regions of a NOESY spectrum acquired in B(80–110) at 5°C. The aromatic resonances of Trp94 are indicated, and medium-range and long-range NOE contacts are labelled. These contacts are indicated in Table 1. The Leu95 Meδ and Ile96 Meδ protons show the same chemical shift. Some of the indicated signals probably involve contacts with both methyl groups. The contact between the 4H proton of Trp94 and Asp93 H_β is observed at a lower contour level.

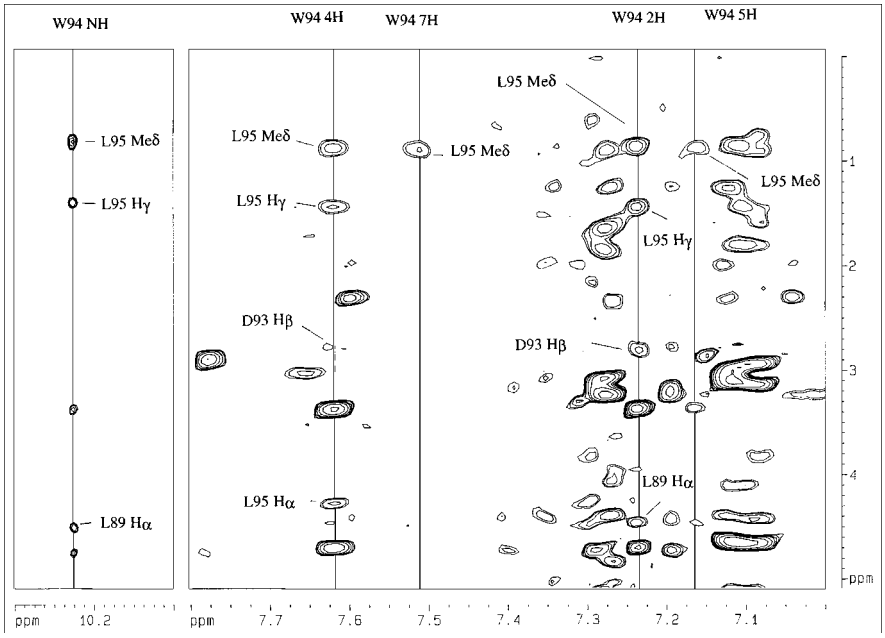
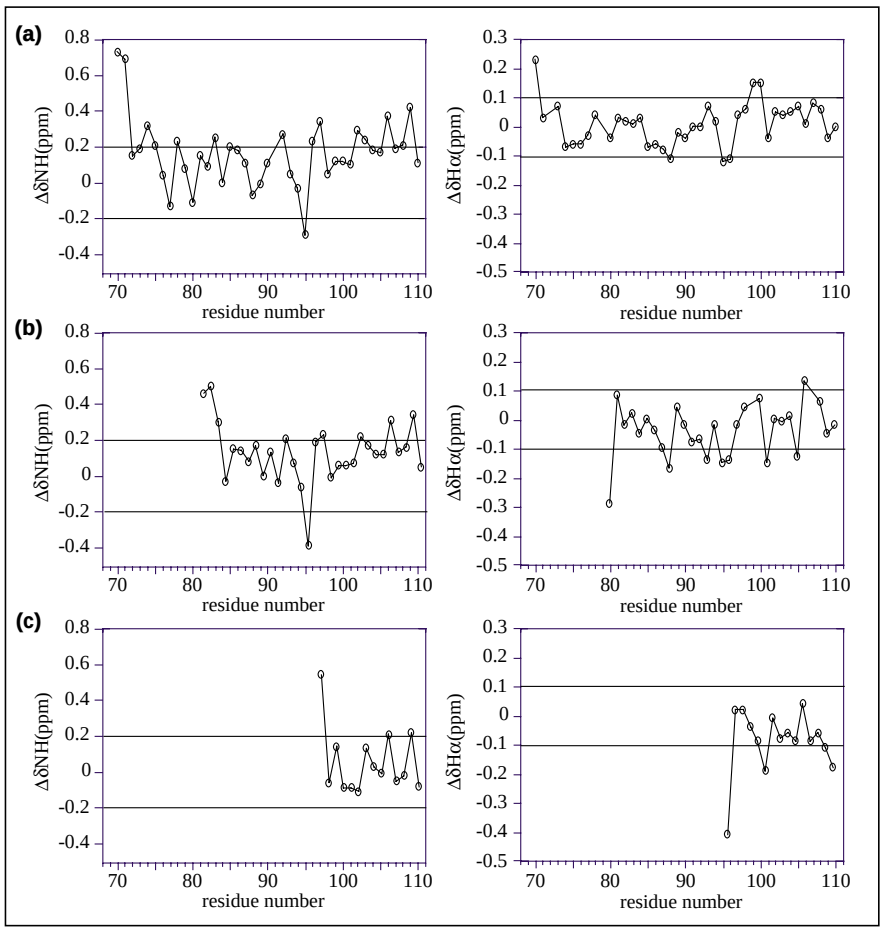


Figure 3

Conformational shifts for (a) B(69–110), (b) B(80–110) and (c) B(96–110). Differences between the chemical shifts of NH in every fragment and the NH of model compounds [26] (left) ($\Delta\delta NH = \Delta\delta NH^{fragment} - \Delta\delta NH^{random\ coil}$). Differences between the H_α protons in every fragment and those of model peptides [26] (right) ($\Delta\delta H_\alpha = \Delta\delta H_\alpha{}^{fragment} - \Delta\delta H_\alpha{}^{random\ coil}$). The lines in the NH and H_α protons indicate the range accepted for the random coil. The H_α protons of Thr99 and Thr107 in B(80–110) were tentatively assigned and they are not represented. In B(69–110), the H_α protons of Arg72 and Thr79, and the NH of Ser91, were tentatively assigned and they are not represented. Amino acids have been numbered according to the intact protein.



tions can render small changes in chemical shifts. Also, the differences can be attributed to changes in the conformational preferences of the aromatic chains (Tyr97, His102, Tyr103 and Phe106) in the fragments. More difficult to explain is the behaviour of Asp93, for which the H_α conformational shift in B(69–110) is close to zero. Probably this results from an altered position of the aromatic sidechain of Trp94 (see Discussion). The cluster of negative deviations in both fragments for the H_α protons of residues Ile88, Asp93, Leu95 and Ile96 provides initial evidence of the presence of stable structured conformations (Fig. 3a,b).

The NH protons are very sensitive to any change in environment [25], and it is not clear whether they are a reliable indicator of residual structure. However, deviations from the accepted random-coil values are observed in the C-terminal region and residues Ser92, Leu95, Ile96 and Tyr97 in both fragments. Also, in B(69–110) deviations are observed in the NH protons of Thr78 and Phe82. Interestingly, these residues are involved in NN(*i*, *i*+1) contacts (see below).

A comparison between the NH protons of common residues in both fragments indicates minor differences in region Ile88–Tyr97. They may result from the same effect described for the H_α proton of Asp93.

Comparison with the chemical shifts of urea-denatured barnase. The chemical shifts of H_α protons for common residues in the urea-denatured state of barnase and fragments are the same, suggesting a similar chemical environment (Fig. 4a,b). Only the H_α proton of Gln104 appears to be downfield shifted. This discrepancy may be due to preferred orientations of the aromatic ring of Tyr103 in the fragments.

There are systematic differences in chemical shifts of the NH protons of both fragments when compared with those of the urea-denatured barnase (Fig. 4a,b). Only those of Ile88, Ile96 and Tyr97 deviate significantly from the rest, although the chemical shift of Ile88 is similar under these conditions to that in denatured barnase. The differences result from different urea concentrations being used (5.5 M for denatured barnase while 6 M urea for the fragments) at different temperatures (30° and 5°C, respectively). The chemical shifts of amide protons are known to change with the temperature, and this change is represented by the temperature coefficients [28,29]. The chemical shifts extrapolated at 30°C from the temperature coefficients obtained (see next paragraph) are in good agreement with those observed in the denatured intact protein for Tyr97 [3] (data not shown). Differences in the chemical shifts of Ile88 and Ile96 from those in denatured barnase may result from different conformations of the aromatic rings.

Temperature coefficient analysis

The temperature coefficients are in the range expected for model compounds, indicating an absence of solvent protection for amide protons (Figs 5,6). The smaller temperature coefficient observed for Trp94 may be explained by burial by the bulky hydrophobic sidechain, as has been found in model compounds [29][1].

NOESY analysis

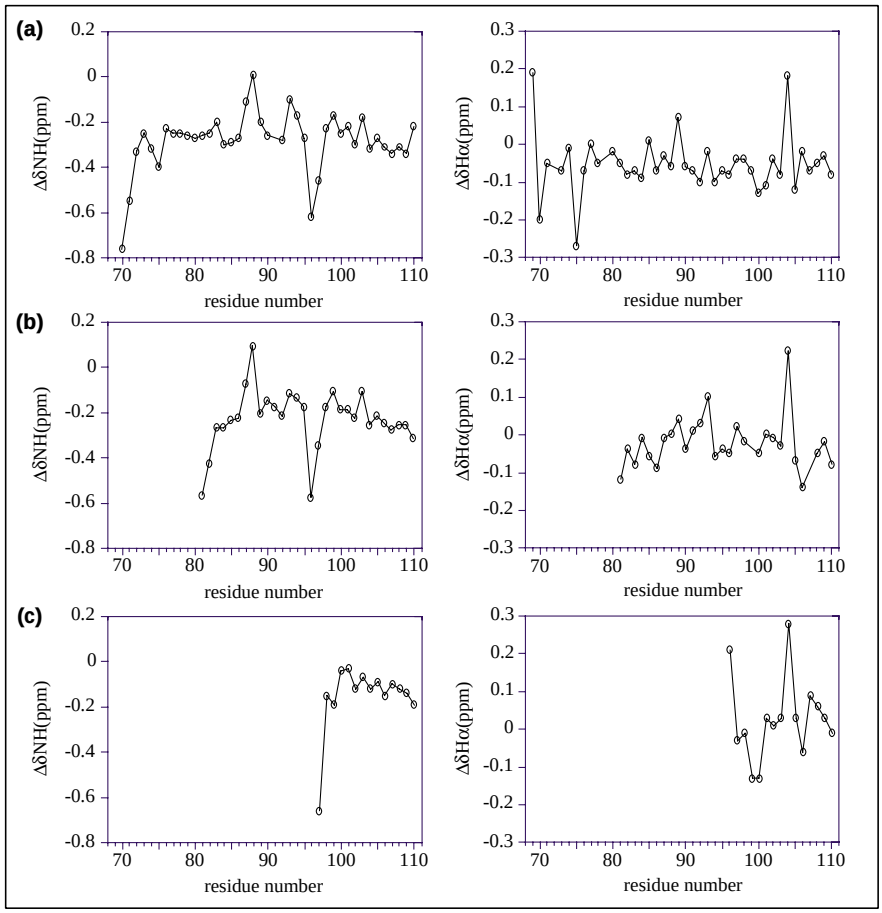
The NOESY spectra of both fragments are dominated by the presence of strong sequential αN(*i*, *i*+1) contacts, typical of random-coil conformations (Figs 5,6). In both fragments, however, there are NOEs that are inconsistent with random-coil conformations. In B(80–110), a strong sequential NN(*i*, *i*+1) NOE between Asp93 and Trp94 is observed (Fig. 6). Also in B(69–110), medium sequential NN(*i*, *i*+1) NOEs are detected in the polypeptide patches comprising residues Asp75–Asn77, Gly81–Arg83 and Asp93–Ile96 (Fig. 5). The NOEs involving residues located in $loop_4$, i.e. those between Gly81 and Arg83, are not observed in B(80–110) because of the proximity to the N terminus.

In B(80–110), the finding of one sequential NN(*i*, *i*+1) NOE in the hairpin region, where deviations in the conformational shifts are observed, indicates the presence of local non-random conformations, probably β-turn-like or α-helix-like [30]. Further, the presence of a weak αβ(*i*, *i*+3) NOE contact between Leu95 and Ser92 suggests the existence of α-helix-like conformations (Fig. 6). Trp94 can be used as a probe to check the presence of any non-random conformation in the turn region, as all of its aromatic protons are well defined, and the indole NH is separated from the rest of the aromatic protons. Non-native long-range and medium-range contacts are observed with the sidechain of Trp94 (Fig. 2; Table 1). Although, aromatic–aliphatic NOE contacts are observed with the rest of the aromatic sidechains in the fragment, the aromatic moiety of Trp94 is the only one involved in long-range contacts (such as those with the H_α of Leu89; Fig. 2).

In B(69–110), the long-range contacts with the indole signals of Trp94 are not observed, only those with the

[1]Wright and co-workers [29] have suggested that there is some degree of conformational preference for the orientation of aromatic residues even in model random-coil peptides. Although this effect remains unclear, it results from an intrinsic property of the sequence, as a consequence of the bulky aromatic chain. According to their suggestions, the low temperature coefficient can be explained because of this intrinsic sequence property. Also, they suggest that the conformational shifts of H_α protons of the two residues either side of the aromatic sidechain should be corrected. Taking into account these corrections, the H_α chemical shifts of residues Ile88 and Leu95 would still be outside of the accepted random-coil value, and those of Asp93 and Ile96 would be below, but very close to the range limit.

Figure 4



Comparison of chemical shifts in denatured barnase and its fragments, (a) B(69–110), (b) B(80–110) and (c) B(96–110). Differences between the chemical shifts of NH in every fragment and the NH of urea-denatured barnase [3] (left) ($\Delta\delta NH = \Delta\delta NH^{urea\text{-}barnase} - \Delta\delta NH^{fragment}$). Differences between the H_α protons in every fragment and those of urea-denatured barnase [3] (right) ($\Delta\delta H_\alpha = \Delta\delta H_\alpha{}^{urea\text{-}barnase} - \Delta\delta H_\alpha{}^{fragment}$). The H_α protons of Thr99 and Thr107 in B(80–110) were tentatively assigned and they are not represented. In B(69–110), the H_α protons of Arg72 and Thr79, and the NH of Ser91 were tentatively assigned and they are not represented. Amino acids have been numbered according to the intact protein.

sidechain of Leu95 remain (Fig. 5). In the intact protein, the aromatic sidechain of Trp94 is involved in tertiary contacts with α-helix$_1$ [31,32] (Fig. 1), but no other contacts are observed with residues involved in the β-turn region. Thus, the presence of the contacts with Trp94 in both fragments indicates the existence of non-native structure, although small changes in the conformational preferences of the aromatic chain of Trp94 are observed in B(69–110).

Further, the rest of the aromatic residues show non-native and medium-range contacts with other hydrophobic sidechains (Figs 5,6), but essentially the same medium-range and aromatic–aliphatic contacts are observed in both fragments.

B(96–110) fragment

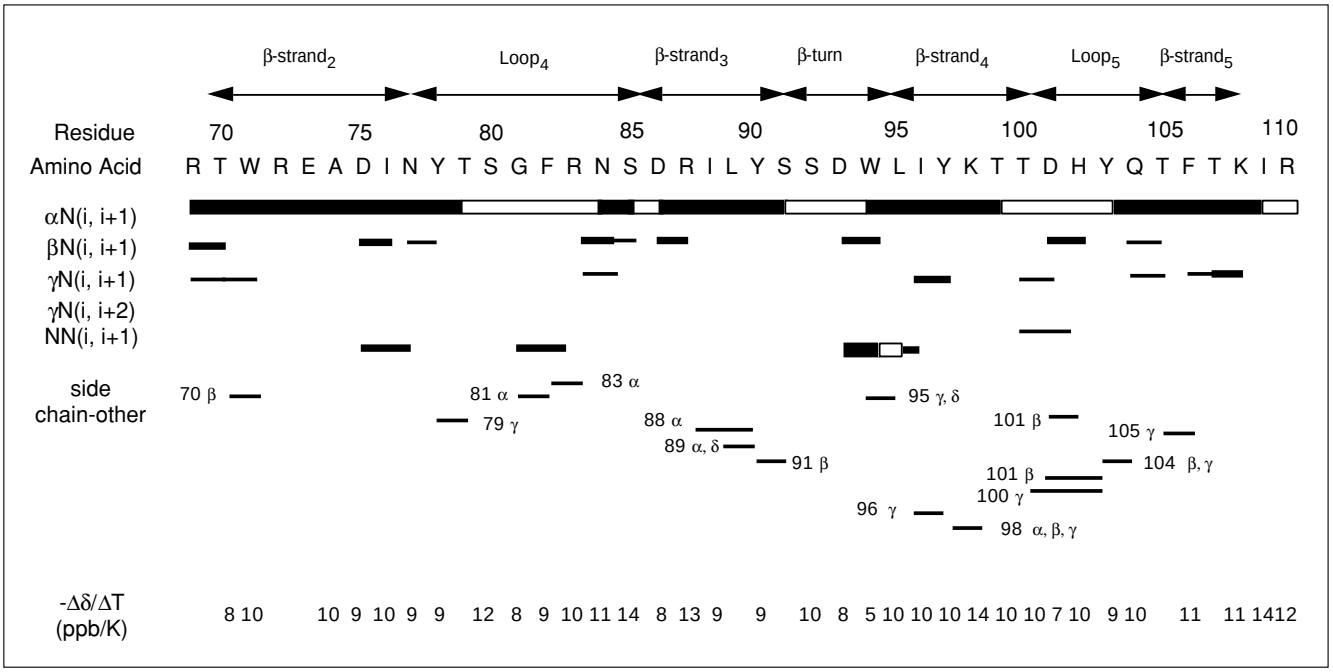
B(96–110) contains β-strand$_4$, loop$_5$, β-strand$_5$ and the last two residues. It is the only fragment that is soluble in aqueous buffer and that does not contain the β-turn region. It was included in this study for two reasons: firstly, to check the presence of non-native hydrophobic contacts with Tyr97, His102, Tyr103 and Phe106 in larger fragments, and secondly, to check the anomalous value of Gln104 when comparison with assignment of the denatured state of the intact protein with the larger fragments was carried out. Sample aggregation was tested by measuring the linewidths and chemical shifts in 1D-NMR. The assignments of B(96–110) in aqueous buffer are provided as Supplementary material.

Chemical shift analysis

Only the H_α proton of Asp101 is slightly upfield shifted, probably from the proximity of His102 (Fig. 3c). This chemical shift was observed in 6 M urea (data not shown) and a similar value was reported in the chemical shift analysis of the denatured state of the intact protein [3]. Also, the NH protons of Phe106 and Ile109 are slightly downfield shifted.

Comparisons with the H_α chemical shifts of larger fragments show excellent agreement (Fig. 4), indicating the similar chemical environment for backbone protons. Because of the sensitivity of NH protons to solvent conditions, there are differences between those of B(96–110) and the other two fragments.

Figure 5



NOE connectivities for B(69–110). Sequence and summary of NOE connectivities and temperature coefficients for the amide protons (bottom) observed in 6 M urea for B(69–110). The thickness of the lines reflects the intensity of the sequential NOE connectivities, i.e. weak, medium and strong. The α(N, β, γ)N(β, γ)(*i, i+j*) indicates an observed contact between the Hα(N, β, γ) proton of a residue and the NH(β, γ) proton of the *i+j* residue. The 'sidechain-other' label indicates those NOE contacts between the aromatic sidechains and a proton of another residue, the name of which is indicated at the end of the line. An open rectangle indicates NOE contacts that are not observed because of closeness to the spectrum diagonal, signal overlapping or overlapping with water signal. The NOE contacts indicated for the sidechain of Trp94 are also in Table 1. The uncertainty in temperature coefficients is ± 1 ppb K^{-1}. The elements of secondary structure are indicated at the top. Amino acids have been numbered according to the intact protein.

Comparison with the chemical shifts of urea-denatured barnase. The chemical shifts of H_α protons of B(96–110) are very close to those observed in the urea-denatured state of barnase [3] (Fig. 4c), except that of Gln104. The chemical shift of this H_α proton is similar to those observed in B(80–110) and B(69–110), as found when B(96–110) is studied in the presence of 6 M urea (data not shown). The differences with NH protons of urea-denatured barnase result from the different temperature and solvent conditions (Fig. 4c). Those differences are not observed when B(96–110) is studied in the presence of 6 M urea (data not shown).

Temperature coefficient analysis

The temperature coefficients are similar to those reported for model compounds [28,29], thus indicating no solvent protection.

NOESY analysis

The NOESY experiment is dominated by strong sequential αN(i, i+1) NOEs typical of extended conformations (Fig. 7). There are non-native hydrophobic contacts with the aromatic sidechains of Tyr97 and Tyr103, although less intense than in the larger fragments (the NOE contacts between the aromatic chain of Tyr103 and the H_β of Asp101 and those with the sidechain of His102 are too weak to be unambiguously identified). Thus, B(96–110) is mainly unstructured in water. Similar results were observed in 6 M urea (data not shown).

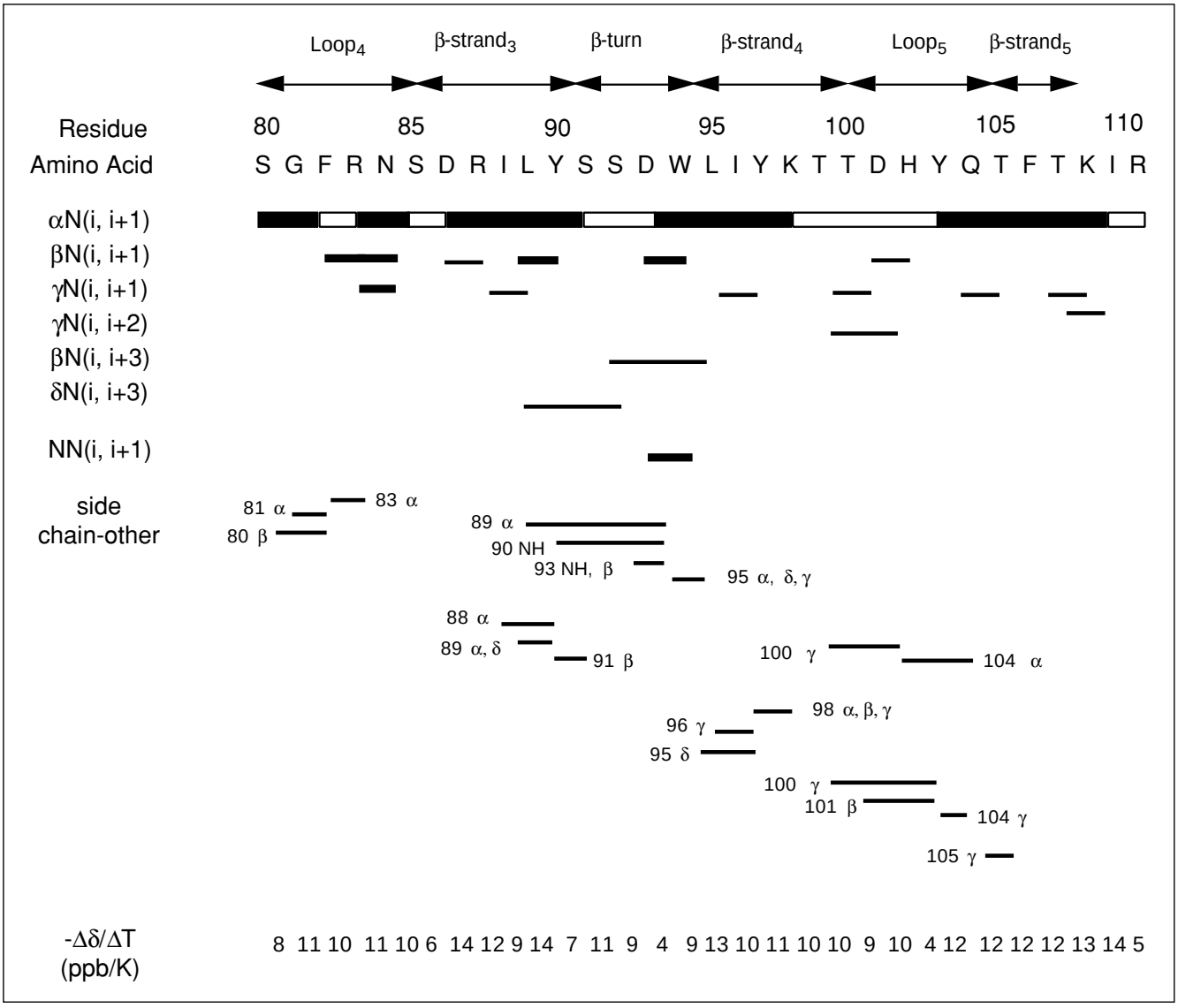
In conclusion, the results observed in B(96–110) indicate that: (i) the conformational preferences around Tyr97, His102, Tyr103 and Phe106 are not length-dependent and they are due to local preferred conformations, and (ii) the observed chemical shift for the H_α proton of Gln104 in larger fragments is similar to that measured in B(96–110), suggesting a preferred conformation of the polypeptide chain in the three fragments.

Discussion

Comparison with urea-denatured state of barnase

The urea-denatured and pH-denatured states of barnase have been assigned and compared [2,3]. In both denatured states, regions with weakly formed residual structure were ascribed to $\alpha\text{-helix}_1$ and the last β-hairpin. In the last β-hairpin, however, assignments were ambiguous because

Figure 6



NOE connectivities for B(80–110). Sequence and summary of NOE connectivities and temperature coefficients for the amide protons (bottom) observed in 6 M urea for B(80–110). The thickness of the lines reflects the intensity of the sequential NOE connectivities, i.e. weak, medium and strong. The α(N, β, γ)N(β, γ)(*i*, *i+j*) indicates an observed contact between the Hα(N, β, γ) proton of a residue and the NH(β, γ) proton of the *i+j* residue. The 'sidechain-other' label indicates those NOE contacts between the aromatic sidechains and a proton of another residue, the name of which is indicated at the end of the line. An open rectangle indicates NOE contacts which are not observed because of closeness to the spectrum diagonal, signal overlapping or overlapping with water signal. The NOE contacts indicated for the sidechain of Trp94 are in Table 1 also. The uncertainty in temperature coefficients is ± 1 ppb K^{-1}. The elements of secondary structure are indicated at the top. Amino acids have been numbered according to the intact protein.

of overlap of signals. With fragments, we can solve this ambiguity, as there are fewer resonances than in the intact protein.

The β-hairpin region is close to α-helix$_1$ in barnase (Fig. 1), forming the major hydrophobic core [31,32]. No interactions between the regions were observed in the pH-denatured and urea-denatured states of the intact protein [2,3]. Further, there are no tertiary interactions observed at all in the solvent-denatured state of barnase. As there are no tertiary interactions in the fragments and solvent conditions are similar to those used in intact barnase, we would not expect any significant differences in the chemical shifts of the fragments and those of the

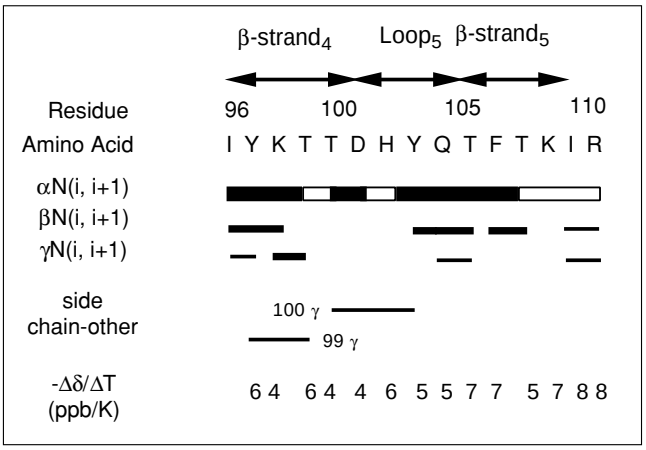
Table 1

Medium-range and long-range NOE interactions with the sidechain of Trp94 in 6 M urea in B(80–110) and B(69–110) – the Leu95 Meδ and Ile96 Meδ protons show the same chemical shift, and probably some of the indicated signals result from contacts with both methyl groups.

Aromatic proton*	NOE contacts†	
	B(80–110)	B(69–110)
NH	Leu89 H_α (m)	–‡
	Tyr90 NH (w)	
	Asp93 NH (w)	
	Leu95 H_γ (m)	
	Leu95 Meδ (s)	
2H (6H)§	Leu89 H_α (w)	
	Asp93 H_β (m)	
	Leu95 H_γ (m)	Leu95 H_γ (w)
	Leu95 Meδ (m)	Leu95 Meδ (w)
4H	Asp93 H_β (w)	
	Leu95 H_α (m)	
	Leu95 H_γ (m)	Leu95 H_γ (w)
	Leu95 Meδ (s)	Leu95 Meδ (w)
5H	Leu95 Meδ (m)	Leu95 Meδ (w)
7H	Leu95 Meδ (m)	Leu95 Meδ (m)

*The nomenclature of the protons of the aromatic moiety was according to Wüthrich [26]. †Intensity of the NOE contacts was judged by visual inspection, counting the number of contour levels in the NOESY spectrum, according to strong (s), medium (m) and weak (w). The contacts with the own backbone protons of Trp94 are not indicated. ‡The NH indole proton does not show any contact in B(69–110). §The 2H and 6H protons show the same chemical shift.

Figure 7



NOE connectivities for B(96–110). Sequence and summary of NOE connectivities and temperature coefficients for the amide protons (bottom) observed in aqueous solution for B(96–110). The thickness of the lines reflects the intensity of the sequential NOE connectivities, i.e. weak, medium and strong. The α(N, β, γ)N(β, γ)(*i*, *i*+*j*) indicates an observed contact between the Hα(N, β, γ) proton of a residue and the NH(β, γ) proton of the *i*+*j* residue. The 'sidechain-other' label indicates those NOE contacts between the aromatic sidechains and a proton of another residue, the name of which is indicated at the end of the line. An open rectangle indicates NOE contacts which are not observed because of closeness to the spectrum diagonal, signal overlapping or overlapping with water signal. The uncertainty in temperature coefficients is ± 1 ppb K^{-1}. The elements of secondary structure are indicated at the top. Amino acids have been numbered according to the intact protein.

intact protein, except from the different temperature used. This has been shown in the H_α conformational shift study for B(69–110) and B(80–110) (except for Gln104).

The NOESY spectra of the fragments are similar to the NOESY spectrum of urea-denatured barnase. In urea-denatured barnase, however, only one NN(*i*, *i*+1) contact is observed between residues Ile96 and Tyr97. In contrast, in the pH-denatured state [3], sequential NN(*i*, *i*+1) contacts are observed in regions Tyr78–Phe82 and Tyr90–Tyr97, i.e. similar to the regions found in B(69–110) in urea. The higher temperature in experiments on denatured barnase, and thus higher mobility, may hamper observation of sequential NN(*i*, *i*+1) NOEs in the intact protein. They could still be present at lower temperature, where populated structures are more stable. Further, the fragments may represent only a small fraction of the total possible conformations of the denatured state. On the other hand, the observed aromatic sidechain contacts with Trp94, which were not assigned in barnase, can be now assigned to the methylene and methyl protons of Leu95 and Ile96, assuming that chemical shifts and NOE contacts are similar in fragments and urea-denatured protein.

Comparison with theoretical simulations of barnase folding

Simulations of the folding of the β-hairpin region comprising residues Ile88–Lys98 have been made — based on the burial of non-polar area, mainchain hydrogen bonds and local mainchain electrostatics — as being the most important factors [33]. Non-native contacts were observed between the sidechains during simulations, and there is a hydrophobic patch predicted to be on one side of the hairpin formed by the sidechains of Ile88, Tyr90, Trp94 and Ile96 [33]. In our experiments, we observe anomalous conformational shifts in the H_α protons of Ile88 and residues clustered around Trp94 (Fig. 3a,b) and also non-native hydrophobic contacts with the other aromatic sidechains (Figs 5,6,7). Further, there is a medium-range NOE between the sidechain of Tyr90 and the H_α of Ile88 (and that of Leu89) in B(80–110) (Fig. 6) and weaker in B(69–110) (Fig. 5). Thus, Ile88, Tyr90, Trp94 and Ile96 residues can get close enough in the fragments to form a

hydrophobic cluster similar to that found in folding simulations.

Implications for protein folding

Turns are observed as intermediates in α-helix/coil transitions [34,35] on a time scale of nanoseconds or less [36]. So far, the study of β-turns and β-hairpins has been restricted to model peptides because the more interesting polypeptide chains have been very insoluble. Recently, however, a native-like β-hairpin conformation was found in an 18-residue fragment of protein G [37] and evidence of β-hairpin-like conformation was also found for this fragment in 6 M urea. These findings suggest that β-hairpins or β-turns could play a role in the folding reaction during the first events.

The region comprising the β-turn in barnase (residues Ser92–Leu95) is well formed in its folding intermediate [17]. It has been predicted to constitute an initiation site for folding based on the hydrophobic surface buried upon folding. Further, the chemical nature of two residues of this region (Ser92 and Asp93) is conserved among other members of the barnase family [38].

We have studied two fragments, B(69–110) and B(80–110), containing the last β-hairpin of barnase; although the polypeptide sequences that are in common have essentially the same chemical shifts, there are some differences in the observed NOE contacts around the β-hairpin region. The sequential backbone NN(i, i+1) and αβ(i, i+3) contacts in B(80–110) involving residues Ser92–Ile96 indicate that the backbone is adopting β-turn-like or α-helix-like conformations (Fig. 6). We suggest that these conformations are stabilized in B(80–110) by the contacts between the sidechains of Trp94 and residues involved in the β-turn region. Also, the involvement of tryptophan residues in local non-random conformations has been described in the folding pathway of basic pancreatic trypsin inhibitor [39,40] and chymotrypsin inhibitor 2 [21,41]. On the other hand, we do not observe all the medium-range and long-range NOEs with the aromatic sidechain of Trp94 in B(69–110), probably because of different conformations being occupied by the aromatic moiety (Fig. 5). The presence of the β-strand$_2$ in B(69–110) may shift the position of the aromatic ring of Trp94 towards more native-like positions, although the same β-turn conformation is maintained as judged from the presence of similar NN(i, i+1) contacts. This is consistent with, firstly, the H_α proton of Asp93 in B(69–110) being closer to random-coil values than that of the Asp93 in B(80–110) and, secondly, the deviations of the chemical shifts of the NH protons of residues Ile88–Tyr97 in B(69–110).

Non-native aromatic–aliphatic contacts with the rest of the aromatic residues (Phe82, Tyr90, Tyr97, His102, Tyr103 and Phe106) are observed in both fragments, indicating that those hydrophobic contacts may be important for local fragment preferences. The aromatic–aliphatic contacts have been observed in other protein fragments and their importance in stabilization of regular secondary structure has been addressed [42].

The presence of non-native NN(i, i+1) NOEs in the edge of β-strand$_2$ and loop$_4$ also indicates that there are some secondary structural preferences in these regions. These regions do not form until late in the folding reaction of barnase [19]. Thus, the sequential preferences observed here, and in the denatured state of barnase [3], are not productive for subsequent folding reaction. Also, similar non-native NN(i, i+1) contacts in residues involved in β-strands have been observed in intact basic pancreatic trypsin inhibitor under denaturing conditions [43,44]. Probably, these non-native interactions, as those described with the aromatic sidechains, reflect the global search through the conformational space during the first folding events.

All these findings can be analyzed according to the nucleation-condensation mechanism proposed recently in folding of chymotrypsin inhibitor 2 [10,11] and in theoretical lattice simulations [12]. According to this model, a nucleation site will develop in the transition state if it makes enough tertiary contacts. Nucleation sites can be unequivocally assigned by using both kinetic and structural analysis, the protein engineering method for the former and NMR for fine analysis of the latter. The β-turn region and α-helix$_1$ in barnase bury a large hydrophobic surface upon folding. Both regions are well formed in the transition state of barnase [16], but there is very little structure in short fragments containing α-helix$_1$ (less than 10%) [20,21]. Here, we observe no highly populated conformations in the β-hairpin region, but non-native-like hydrophobic clustering is observed around Ile88, Asp93, Trp94, Leu95 and Ile96. Probably, these non-native interactions rearrange when docking of this region with α-helix$_1$ occur, i.e. when additional long-range interactions are made. In contrast, the region forming β-strand$_2$ and loop$_4$ does not make extensive tertiary contacts. This precludes its consolidation as a nucleation site and its formation until late in the folding pathway. Thus, although in β-strand$_2$ and loop$_4$ there are secondary structural preferences (as judged from the sequential NN(i, i+1) contacts), they do not constitute a nucleation site.

In conclusion, it appears that the β-hairpin region in all fragments analyzed is only weakly structured in early stages of folding. Similar results have been found in α-helix$_1$ (the other proposed 'nucleation site' in barnase), although the structure seems to be native-like [20,21]. In contrast, the structure in the β-hairpin region is non-native, but it is consistent with the suggestion that this

turn could nucleate the β-sheet fold. The non-native character of the structure could prevent aggregation and it could reflect the conformational search in the first events of folding reaction.

Materials and methods

Chemicals and buffers

Cyanogen bromide was purchased from Fluka. D_2O was purchased from Fluorochem Ltd. TSP was from Aldrich, and d_3-acetic, and d_3-sodium acetate were from SIGMA. All other reagents were of analytical grade and water was deionized and purified on an Elgastat system.

Mutagenesis and protein purification

Site-directed mutagenesis was performed as described previously [20]. The mutants used here were Gly68→Met, Trp79→Met and Leu95→Met, with methionine at positions 68, 79 and 95, respectively. The positions of the mutation were specifically chosen at points between elements of secondary structure that are largely exposed to the solvent and make few contacts with other sidechains. Mutant proteins were purified from the supernatant of an *Escherichia coli* culture as described for wild-type enzyme [45]. Purified proteins were flash frozen and stored at –70°C.

Cleavage reaction

Lyophilized protein was dissolved in water at a final concentration of 1 mg ml^{-1}. The pH of the solution was adjusted to 1.5 with HCl. Cyanogen bromide was dissolved to a concentration of 3–5 mg per mg of protein. The reaction was allowed to proceed at room temperature in the dark and was stopped after 8–10 h by adding four volumes of distilled water to the mixture. The mixture was frozen in liquid nitrogen and lyophilized to remove the excess of reagents.

Purification of fragments

The lyophilized mixture cyanogen bromide cleavage reaction was purified using a Gilson HPLC apparatus with a Rainin-Dynamax reversed phase semipreparative C_8 column and an acetonitrile gradient (0.1% TFA) with absorbance monitored at 280 nm. The molecular weight of peptides was confirmed by matrix-assisted laser desorption time-of-flight mass spectrometry. After purification, fragments B(69–110) and B(80–110) were not soluble in aqueous buffer. Peptide concentrations were determined from the extinction coefficient of model compounds [46]. Purified peptides were lyophilized and stored at –20°C.

^1H nuclear magnetic resonance spectroscopy

Lyophilized protein fragments were dissolved in 0.5 ml 50 mM buffer (sodium d_3-acetate, pH 4.5) in 90% H_2O/10% D_2O or in the presence of 6 M urea for B(69–110) and B(80–110). An 8 M urea stock solution was prepared and the corresponding amounts were added to the NMR samples to give a final concentration of 6 M urea. The pH was measured using a glass Russell electrode. No correction was made for isotope effects. In the presence of urea, the pH was adjusted with small amounts of NaOH or HCl to give a final value of 4.5. Fragment concentrations were in the range 1–1.5 mM for B(96–110) and 2–2.5 mM for B(69–110) and B(80–110).

The spectra were recorded on a Bruker AMX-500 spectrometer at 5°C. 2D-NMR experiments were acquired in the phase-sensitive mode using the TPPI method [47] with presaturation of the water signal during relaxation delay. TOCSY [48,49], NOESY [50] and ROESY [51] were recorded using standard phase-cycling sequences. Usually, spectra were acquired with 2K data points in t_2 and 256 t_1 increments with a spectral width of 8000 Hz in both dimensions. Typically, 128 scans were collected per t_1 increment. In TOCSY experiments, a 70 ms mixing time, and 150 ms (and 125 ms) and 125 ms for NOESY and ROESY experiments, respectively, were used. The temperature dependence of amide signals of every fragment was followed by TOCSY experiments at several temperatures within the 5–20°C range. The temperature coefficients for each amide proton were obtained from the slope of the least-square fit to a straight line. Also, a NOESY experiment was acquired at 10°C to avoid overlapping of the signals at 5°C.

Spectra were processed using BRUKER-UXNMR software on an Aspect X-32 workstation. Prior to Fourier transformation, data matrix was zero-filled to obtain a final matrix size of 2K × 512 words. The square sine bell window functions, shifted $\pi/4$ in both dimensions, were applied for all spectra. The phase shift was optimized for every spectrum and polynomial base-line correction was applied in all cases in both dimensions. ^1H chemical shifts are quoted relative to external TSP (0.0 ppm), under the same conditions used in the experiments.

The complete assignment of the ^1H NMR spectra of each peptide in aqueous and urea solutions was performed by using the sequence-specific method [26].

Supplementary material

^1H assignments are available as Supplementary material published with this paper on the internet.

Acknowledgements

JL Neira was supported by an EC fellowship.

References

1. Fersht, A.R. (1993). Protein folding and stability: the pathway of folding of barnase. *FEBS Lett.* 325, 5–16.
2. Arcus, V.L., Vuilleumier, S., Freund, S.M.V., Bycroft, M. & Fersht, A.R. (1994). Toward solving the folding pathway of barnase: the complete backbone ^{13}C, ^{15}N and ^1H NMR assignments of its pH-denatured state. *Proc. Natl. Acad. Sci. USA* 91, 9412–9416.
3. Arcus, V.L., Vuilleumier, S., Freund, S.M.V., Bycroft, M. & Fersht, A.R. (1995). A comparison of the pH, urea and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J. Mol. Biol.* 254, 305–321.
4. Baldwin, R.L. (1993). Pulsed H/D exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* 3, 84–91.
5. Fersht, A.R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224, 771–782.
6. Fersht, A.R. (1995). Characterising transition-states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* 5, 79–84.
7. López-Hernández, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Folding & Design* 1, 43–55.
8. Wetlaufer, D.B. (1990). Nucleation in protein folding. Confusion of structure and process. *Trends Biochem. Sci.* 15, 414–415.
9. Anfinsen, C.B. & Scheraga, H.A. (1975). Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* 29, 205–300.
10. Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254, 260–288.
11. Fersht, A.R. (1995). Optimisation of rates of protein folding: the nucleation-collapse mechanism and its implications. *Proc. Natl. Acad. Sci. USA* 92, 10869–10873.
12. Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. (1994). Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33, 10026–10036.
13. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* 29, 7133–7155.
14. Zefhus, M. (1995). Automatic recognition of hydrophobic clusters and their correlation with protein folding units. *Protein Sci.* 4, 1188–1202.
15. Shortle, D. (1996). Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* 6, 24–30.
16. Serrano, L., Matouschek, A. & Fersht, A.R. (1992). The folding of an enzyme III. Structure of the transition state for the unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* 224, 805–818.
17. Matouschek, A., Serrano, L. & Fersht, A.R. (1992). The folding of an enzyme IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* 224,

819–835.
18. Matouschek, A., Serrano, L., Meiering, E.M., Bycroft, M. & Fersht, A.R. (1992). The folding of an enzyme V. H/^{2}H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J. Mol. Biol.* 224, 837–845.
19. Serrano, L., Matouschek, A. & Fersht, A.R. (1992). The folding of an enzyme VI. The folding pathway of barnase: comparison with theoretical models. *J. Mol. Biol.* 224, 847–859.
20. Sancho, J., Neira, J.L. & Fersht, A.R. (1992). An N-terminal fragment of barnase has residual helical structure similar to that in a refolding intermediate. *J. Mol. Biol.* 224, 749–758.
21. Kippen, A.D., Arcus, V.L. & Fersht, A.R. (1994). Structural studies corresponding to mutants of the major α-helix of barnase. *Biochemistry* 33, 10013–10021.
22. De Prat Gay, G., Ruiz-Sanz, J., Davis, B. & Fersht, A.R. (1994). The structure of the transition state for the association of two fragments of the barley chymotrypsin inhibitor 2 to generate native-like protein: implications for mechanism of protein folding. *Proc. Natl. Acad. Sci USA* 91, 10943–10946.
23. De Prat Gay, G., Ruiz-Sanz, J., Neira, J.L., Itzhaki, L.S. & Fersht, A.R. (1995). Folding of a nascent polypeptide chain *in vitro*: cooperative formation of structure in a protein module. *Proc. Natl. Acad. Sci USA* 91, 10943–10946.
24. Wishart, D.S., Sykes, B.D. & Richards, F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* 222, 311–333.
25. Oldfield, E. (1995). Chemical shifts and three-dimensional protein structures. *J. Biomol. NMR* 5, 217–225.
26. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, Inc., New York.
27. Neri, D., Wider, G. & Wüthrich, K. (1992). Complete ^{15}N and ^{1}H proton assignments for the amino terminal domain of the phage 434 repressor in the urea unfolded form. *Proc. Natl. Acad. Sci USA* 89, 4397–4401.
28. Jiménez, M.A., Nieto, J.L., Rico, M., Santoro, J., Herranz, J. & Bermejo, F.J. (1986). A study of the NH NMR signals of Gly-Gly-X-Ala tetrapeptides in H_2O at low temperature. *J. Mol. Struct.* 143, 435–438.
29. Merutka, G., Dyson, H.J. & Wright, P.E. (1995). 'Random coil' ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series. *J. Biomol. NMR* 5, 14–24.
30. Dyson, H.J. & Wright, P.E. (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* 20, 519–538.
31. Bycroft, M., Shepard, R.N., Lau, F.T.-K & Fersht, A.R. (1990). Sequential assignment of the ^{1}H nuclear magnetic resonance spectra of barnase. *Biochemistry* 29, 7425–7432.
32. Bycroft, M., Ludvigsen, S., Fersht, A.R. & Poulsen, F.M. (1991). Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry* 30, 8697–8701.
33. Avblej, F. & Moult, J. (1995). Determination of the conformation of folding initiation sites in proteins by computer simulation. *Proteins* 23, 129–141.
34. Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. & Wright, P.E. (1988). Folding of inmunogenic peptide fragments of proteins in water solution I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* 201, 161–200.
35. Dyson, H.J., Rance, M., Houghten, R.A., Lerner, R.A. & Wright, P.E. (1988). Folding of inmunogenic peptide fragments of proteins in water solution II. The nascent helix. *J. Mol. Biol.* 201, 201–217.
36. Tobias, D.J., Mertz, E., Brooks, C.L. III (1991). Nanosecond timescale folding dynamics of a pentapeptide in water. *Biochemistry* 30, 6054–6058.
37. Blanco, F.J., Rivas, G.M. & Serrano, L. (1994). A short linear peptide that folds into a native stable β-hairpin in aqueous solution. *Nat. Struct. Biol.* 1, 584–590.
38. Hill, C., *et al.*, & Pavlovsky, S. (1983). The structural and sequence homology of a family of microbial ribonucleases. *Trends Biochem. Sci.* 8, 364–369.
39. Keeming, J. & Creighton, T.E. (1993). Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. *J. Mol. Biol.* 234, 861–878.
40. Keeming, J. & Creighton, T.E. (1995). The physical properties of local interactions of tyrosine residues in peptides and unfolded proteins. *J. Mol. Biol.* 245, 251–260.
41. Itzhaki, L.S., Neira, J.L., Ruiz-Sanz, J., De Prat Gay, G. & Fersht, A.R. (1995). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor-2. *J. Mol. Biol.* 254, 289–304.
42. Muñoz, V., Serrano, L., Jiménez, M.A. & Rico, M. (1995). Structural analysis of the peptides encompassing all α-helices of three α/β parallel proteins: che-Y, flavodoxin and p21-Ras: implications for α-helix stability and the folding of α/β parallel proteins. *J. Mol. Biol.* 247, 648–669.
43. Pan, H., Barbar, E., Barany, G. & Woodward, C. (1995). Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry* 34, 13974–13981.
44. Barbar, E., Barany, G. & Woodward, C. (1995). Unfolded BPTI variants with a single disulphide bond have diminished non-native structure distant from the crosslink. *Folding & Design* 1, 65–76.
45. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A.R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability: investigation by double-mutant cycles. *Biochemistry* 29, 9343–9352.
46. Gill, S.C. & Von Hippel, P.H. (1989). Extinction coefficients from amino acid sequence data. *Analyt. Biochem.* 182, 319–326.
47. Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of proton-proton spin–spin coupling constants. *Biochem. Biophys. Res. Commun.* 11, 967–975.
48. Braunschweiler, L. & Ernst, R.R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* 53, 521–528.
49. Bax, A. & Davis, D.G. (1985). MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 65, 355–360.
50. Jeener, J., Meier, B., Backman, P. & Ernst, R.R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* 71, 4546–4550.
51. Bothner-By, A.A., Stephens, R.L., Lee, J.M., Warren, C.D. & Jeanloz, R.W. (1984). Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* 106, 811–813.
52. Kraulis, P.J. (1991). MOLSCRIPT, a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be found in the BioMedNet library at http://BioMedNet.com/ – for further information, see the explanation on the contents page.